



## INTELLIGENCE ARTIFICIELLE

### BIG DATA ET TRAITEMENT DE GRANDES MASSES DE DONNEES

#### OBJECTIFS

Comprendre et maîtriser les concepts du Big Data. Introduction au Big Data comprehensive. Valeur ajoutée du Big Data Analytics et défis technologiques et organisationnels. Investissement et compétitivité du Big Data Analytics. Architectures et accès aux grandes masses de données.

#### CONTENU PÉDAGOGIQUE

##### INTRODUCTION

- › Caractérisation du big data
- › Enjeux du big data
- › Problèmes d'acquisition, d'hétérogénéité, de qualité, de traitement...

##### Bases de données NoSQL

- › Bases de données clés-valeurs
- › Bases de données orientées colonne
- › Bases de données orientées document
- › Bases de données orientées graphe

##### Extraction d'informations du Web

- › Reconnaissance d'entités nommées
- › Désambiguation
- › Notions de Web sémantique
- › RDF et SPARQL

##### Traitement distribué

- › Intégration de données massives: approches
- › Principe de partitionnement et de la distribution des données, que ce soit dans un contexte (i) de données relationnelles gérées par une ou plusieurs instances de systèmes de gestion de bases de données et (ii) de données 'simplement' stockées dans des fichiers;
- › Problèmes de performance des traitements/requêtes distribués;
- › Paradigme Map/Reduce et de ses dérivés (e.g. Spark);

##### Flux de données

- › Acquisition
- › Modélisation
- › Filtrage
- › Approches d'analyse de flux de données

##### Compétences ciblées

- › Comprendre ce qui se cache derrière le mot BIG DATA
- › Comprendre l'intérêt et les enjeux du BIG DATA
- › Comprendre la différence entre NoSQL et les bases de données structurées
- › Identifier les principales familles et caractéristiques des bases de données NoSQL
- › Choisir la technologie NoSQL adaptée à un contexte
- › Acquérir une expertise sur les technologies MongoDB, Elasticsearch, Redis et Neo4J
- › Comprendre l'intérêt des plateformes de traitements parallélisés (Hadoop, Spark)
- › Comprendre l'écosystème Hadoop
- › Développer des traitements parallélisés sous forme de batch et de micro-batch avec SPARK



#### DURÉE

3 jours  
21 heures



#### SESSIONS

- 19 - 21 mai 2025 en présentiel à Lyon



#### FRAIS D'INSCRIPTION (DÉJEUNER INCLUS)

1 750 € HT



#### PRÉREQUIS & PUBLIC CONCERNÉ

Chef d'entreprise, directeur, responsable de projet, ingénieur, technicien  
Connaissances générales en informatique

CPE Lyon Formation Continue

41 rue Garibaldi - 69006 LYON

[04.72.32.50.60](tel:0472325060)